Gesture-based mobile training of intercultural behavior

Rehm, Matthias; Leichtenstern, Karin

Published in:
Multimedia Systems

DOI (link to publication from Publisher):
10.1007/s00530-011-0239-8

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Multimedia Systems
DOI 10.1007/s00530-011-0239-8

REGULAR PAPER

Gesture-based mobile training of intercultural behavior

Matthias Rehm · Karin Leichtenstern

Abstract Cultural heuristics determine acceptable verbal and non-verbal behavior in interpersonal encounters and are often the main reason for problems in intercultural communication. In this article, we present an approach to intercultural training of non-verbal behaviors that makes use of enculturated virtual agents, i.e. interactive systems that take cultural heuristics for interpreting and generating behavior into account. Because current trends in intercultural training highlight the importance of a coaching approach, i.e. the ability to offer training units anytime and anywhere, the system was developed as a mobile solution taking the sensoric capabilities of smart phones into account for the user interaction in form of gesture recognition. After an introduction of the theoretical background on culture and enculturated systems, the system features are discussed in detail followed by an account of the application itself, emphasizing the importance of situated role-plays. Two evaluation studies are presented next that analyze the usability of the approach as well as the more important question of whether training with the system gives better results than traditional methods.

Keywords Virtual agents · Mobile edutainment · Serious games · Enculturated systems

M. Rehm (✉)
Department of Media Technology, Aalborg University,
Niels Jernes Vej 14, 9220 Aalborg-Ø, Denmark
e-mail: matthias@create.aau.dk

K. Leichtenstern
Human Centered Multimedia, Augsburg University,
Universitätsstr. 6a, 86159 Augsburg, Germany
e-mail: leichtenstern@informatik.uni-augsburg.de

1 Introduction

We are living in a so-called globalized world that seems to make it possible to communicate without boundaries across different cultures and continents, sometimes acquiring the necessary language skills but often relying on English (or what we non-native speakers claim to be English) as the lingua franca. But communication is not only concerned with getting the message across verbally. It is inherently multimodal ranging from communication management like coordination of turn-taking behavior over facial expressions and gestures to spatial behavior, which often follow culturally determined heuristics. As an example, consider a dinner table discussion, for which the structure of this multiparty conversation can vary from a turn after turn sequence to a situation where several interactions and discussions take place at the same time between different participants. Often such nonverbal aspects of communication give rise to severe misunderstandings [40]. For instance, the first group in our example might classify the second one as chaotic and unfocused, whereas the second group might think of the first one as restrained, distant, and cold. Another well-studied example is the use of space in interpersonal encounters [11]. While for instance in Northern Europe a certain distance between interlocutors is generally acceptable, in an Arabic context, this distance should not be to large in order to allow for touching between interlocutors. Again, the interpretation of the other group's behavior is bound to differ, often resulting in the first group finding the second group invasive or pushy and the second thinking about the first as distant and cold. This is due to the fact that behavior is interpreted based on unconscious cultural heuristics that are formed by our personal interaction histories in the cultural groups to which we belong.

Published online: 02 July 2011

When we talk about enculturated interactive systems, we think about computer systems that take these cultural heuristics of behavior into account when structuring the interaction with the user. Enculturated agent systems as a subspecies of interactive systems make use of an embodied interface in the form of a virtual (or recently also physical) agent that can utilize a rich repertoire of communication channels like speech, gaze, facial expressions, gestures, and others. Possible application areas for such enculturated agent systems include (a) information presentation, where agents become more efficient in delivering information or selling a point or a product by adapting their communication style to the culturally dominant persuasion strategy; (b) entertainment, where a game becomes more entertaining by providing coherent behavior modifications for in-game characters based on their cultural background; and (c) education, where experience-based role-plays become possible for increasing cultural awareness of users, e.g. by augmenting the standard language textbook with behavioral learning.

In this article, we are focusing on this last point. In general, virtual agents offer natural interaction possibilities because of their potential to emulate verbal and nonverbal human behavior (e.g. [4]). Virtual characters have also been shown to be engaging tools for tutoring systems and present a good starting point for exemplifying different perspectives in intercultural training. Thus, our first motivation comes from research in virtual agents and recent efforts to enculturate these systems (e.g. [32, 35]).

Current trends in intercultural training emphasize the importance of a coaching approach [9]. Coaching in this context means centered on the trainee's needs and goals, and especially on his agenda resulting in an anytime anywhere approach with small-scale experience-based learning sessions (i.e. role plays) tailored to the specific context and situation. For instance, being at the train station triggers a learning session on how to purchase a train ticket. Or an imminent meeting with your boss in the afternoon triggers a lesson on how to behave towards a person with higher social status, which greatly differs between cultures. The coaching idea is the second motivation for our work, i.e. the possibility to engage in an experience-based training unit anytime anywhere.

In this article we present Gesture-activated mobile edutainment (GAME), an application that bundles our activities in enculturated agent research from the past years (see, e.g. [36] for an overview) and makes it applicable on a mobile platform. It allows training culture-specific gestures making use of the sensor technology of current smartphones and applying the gestures in role-plays with virtual characters. We start with a thorough investigation in the theoretical underpinnings of enculturated agent system (Sect. 2) Afterwards, we introduce the main technical building blocks of the application itself (Sect. 3) before we present the two interaction modes that allow to either train non-verbal behavior directly or apply one's knowledge and skills about these behaviors in role-plays with virtual agents (Sect. 4). The evaluation of the approach (Sect. 5) covers a usability test as well as an in-depth evaluation on the effectiveness of the approach in terms of acquiring culture-specific behaviors. We conclude with an outline of future work based on the results of the evaluation study (Sect. 6)

2 Theoretical background

The GAME approach brings together different research directions from cultural training over role plays with virtual characters to mobile learning in a comprehensive edutainment scenario drawing heavily from previous work in these diverse areas. In the following, a short introduction is given to diverse backgrounds.

2.1 Culture

The notion of enculturated interactive systems entails the need to define what culture is and how it is relevant for an interactive system. In the introduction some examples were given how cultural heuristics influence face-to-face behavior and its interpretation by others. To be able to model such heuristics in a system, the notion of culture has to become a parameter of the system, i.e. it must be brought in an operational form that can be applied to decide for specific system behaviors.

The notion of culture itself is a multiply defined notion that gives rise to many misconceptions ranging from theater and art over language and national affiliation. Thus, it is necessary to specify exactly what is meant by culture in the envisioned training system as this notion affects several levels of the system like the content of the learning scenarios or the behavior of the virtual characters. We claim that it is indispensable to base a system that integrates cultural aspects of interaction on a thorough theoretical foundation that allows for reliably predicting patterns of behavior that are influenced by cultural heuristics. Hofstede [15] presents a starting point with his theory of cultural dimensions that defines culture as a five-dimensional concept and relates positions on the dimensions to certain behavioral heuristics.

Table 1 gives an overview of behavior patterns that according to Hofstede et al. [16] are related to the high and low values on the cultural dimensions. For instance, in collectivistic cultures (low on identity dimension) people tend to speak softer and stand closer together in interpersonal encounters, whereas in individualistic cultures (high

Table 1 Synthetic cultures and corresponding patterns of behavior for low (L) and high (H) values ([33] following [16])

Dimension	Synthetic culture	Sound	Space
Hierarchy	L: Low power	Loud	Close
	H: High power	Soft	Far
Identity	L: Collectivistic	Soft	Close
	H: Individualistic	Loud	Far
Gender	L: Femininity	Soft	Close
	H: Masculinity	Loud	Close
Uncertainty	L: Tolerance	Soft	Close
	H: Avoidance	Loud	Far
Orientation	L: Short-term	Soft	Close
	H: Long-term	Soft	Far

on identity dimension) people tend to do the opposite, i.e. speak louder and stand further apart in interpersonal encounters. The five dimensions have the following meanings:

1. *Hierarchy* describes the degree to which different distribution of power in a culture is accepted by the less powerful members, ranges from low-power distance (power is, e.g. the result of a vote and thus temporary) to high-power distance (power is linked to a person, e.g. by individual charisma)
2. *Identity* describes the degree to which individuals are integrated into groups, ranging from individualistic (loose ties between individuals) to collectivistic (integration in strong, cohesive in-groups)
3. *Gender* describes the distribution of roles between the genders, ranges from feminine (roles do not differ much) to masculine (clear distinction between gender roles)
4. *Uncertainty* describes the tolerance for uncertainty and ambiguity, ranging from tolerance (more comfortable in unstructured and novel situations) to avoidance (uncomfortable in unstructured and novel situations leading to rules for avoiding such situations)
5. *Orientation* distinguishes between long- and short-term orientation, where long-term orientation is associated with thrift and perseverance, whereas short-term orientation is associated with respect for tradition, fulfilling social obligations, and saving one's face.

As Table 1 exemplifies, with the dimensional model it becomes possible to predict behavioral tendencies based on the position of a culture in this five-dimensional space. There are many shortcomings of this theory esp. related to the sample used for empirical analysis. Nonetheless, Hofstede's work has been successfully adapted in the area of cultural usability (e.g. [27, 28]), whereas attempts for enculturating interactive system have so far been mostly ad hoc and often without a thorough theoretical or empirical foundation (a detailed analysis can be found in [32]).

Hofstede's work comes from a school of thought that broadly equates culture with a set of norms and values that constrain thinking and behavior of the members of a given cultural group. Thus, being able to specify the set of norms and values for a given culture in principle allows deriving decision making processes and behavioral patterns for an interactive system. It is an ongoing debate what these norms and values are. Kluckhorn and Strodtbeck [24] name five value orientations including people and nature, time sense, and social relations. But in their approach the impact of these value dimensions on individual behavior is not evident and thus it remains unclear how their approach can be translated into a computational model. A more recent approach by Schwartz and Sagiv [39] defines values as fundamental heuristics of behavior. Those values can be seen as central goals members of a cultural group aim to achieve and they are based on three universal needs, i.e. biological (the need to eat and drink, etc.), coordinated social interaction (the need to interact with others), and group functioning (the need to make social groups work on a relational and task level). Following this approach, cultural differences originate from different goals or from prioritizing different goals. Again, the impact on individual behavior is unclear. Apart from Hofstede, the work of Hall ([10, 11, 12]) is the one most often employed to model culture specific behavior often relying on his analysis of proxemics, i.e. interpersonal spatial behavior. Hall focuses on three different dimensions, space, time, and context, and defines dichotomies on each dimension. Thus, he distinguishes between high- and low-contact cultures for spatial behavior, monochronous and polychronous cultures for time perception, and low- and high-context cultures for group membership and patterns of communication. He also associates behavior patterns with these dichotomies, e.g.

high-contact cultures are those in which people display considerable interpersonal closeness and immediacy.

This short overview shows that cultural theories regarding norms and values present an interesting starting point for modeling culture-specific behaviors in interactive systems as they seem to allow associating behavioral heuristics with the proposed value dimensions. Hall as well as Hofstede give some explicit examples and are the approaches that are currently the most frequently used ones.

2.2 Enculturated interactive systems

The term enculturated interactive systems describes recent attempts of taking cultural aspects of interaction into account for the design as well as the behavior of interactive systems (see [35] for an overview). Many of these attempts are located in the area of intelligent tutoring systems with the aim of intercultural training, i.e. allowing the user to experience and train culture-specific communication behaviors.

The commercially most successful intelligent tutoring system focusing on cultural aspects is the tactical language training [19], which employs virtual characters in role-playing scenarios. It is used as a language training for soldiers that face expatriate missions. In the training sessions, the users have to solve tasks by employing their language knowledge in the given situation. The main interaction modality is speech. Additionally, users can select gestures to accompany their utterances that are then played as an animation of their character. Culture is equated in this case with the language that is trained and used as a back story for creating animations for the virtual characters. The training goal is language proficiency.

In [23] an intelligent tutoring system is described that is tailored at teaching business etiquette in intercultural encounters. Again, culture is used as a back story for the role-play with a virtual character that determines the "production design". The system aims at teaching (stereo-) typical rules of behavior like "do not bring alcohol as a present in Arabic countries", and allows the user to put his knowledge about such rules to a test in a kind of adventure game. The interaction is realized as a text input.

The aforementioned systems focus on language and knowledge about cultural rules. According to Ting-Toomey [40], the most severe misunderstandings in intercultural communication arise due to different perspectives on appropriate non-verbal behavior in communicative situations. A parameter-based model of culture is described in [18], where certain non-verbal behaviors (proxemics, gaze) of virtual agents are modified in a culture-specific way (US, Mexican, and Arabian) relying on the model parameters. The necessary data for their approach are drawn from a literature review. It turns out that the information from the literature is in most cases merely qualitative in nature, often gives only mean values or does not give information about a culture under investigation. A consequence of this is a mix of culture-specific behavior, e.g. American turn-taking with Arabian proxemics and gaze, which makes it difficult to pinpoint effects found in preliminary perception studies to cultural variables. A similar problem was encountered in [36], where a thorough empirical study is presented to deal with the lack of missing data. Based on the results and Hofstede's dimensional model [15], a probabilistic model of non-verbal behavior is derived, which is employed to categorize and interpret observed user behavior and to control the animations of virtual characters. The user can actually perform non-verbal behavior, e.g. by using a Wiimote, which allows for executing and analyzing gestures. Thus, it becomes possible to give the user a direct feedback on his performance. A prototype is described that gives feedback to a user on his performance by adapting the non-verbal behavior of a group of agents. That the collected data presents a rich source for comparative analyses is exemplified in [8], where cultural aspects of verbal interaction are modeled based on an analysis of the data corpus. A plan-based approach for realizing culture-specific small talk between virtual agents in first meetings is developed based on the empirical insights gained from the German and Japanese recordings.

That the neglect of a thorough cultural model can result in quite dubious systems is exemplified in [43] with a collaborative role-playing game. The approach is problematic because the game itself is culturally biased as it is a typical Western military action game, the creation of the two groups that are compared is invalid as they compare US American teams with multinational teams, and possible decisions in the game seem to be solely based on the developers' intuition and thus their own cultural background.

Whereas all of the above systems focus on observable verbal or non-verbal behavior, quite a different cultural influence, i.e. the internal structure of interaction, has been investigated in [22]. Facing the challenge of developing a smoking cessation game for Maori users, an analysis of persuasive strategies revealed that they are tailored to an individualistic audience, whereas the target users come from a collectivistic culture. Thus, a persuasive game is developed that realizes persuasion strategies which take the collectivistic perspective into account.

2.3 Experience-based training of (non-verbal) behavior

All of the above systems make use of virtual characters as a useful tool for training. Isbister [17] has convincingly

argued for the use of agents to further cross-cultural communication skills between users. Compared with life role-playing games, learning with virtual agents offers additional new experiences that can further the learning process.

- Repeatability: The training scenario can be repeated as often as necessary without annoying a human training partner. Moreover, either one user can repeat a given lesson until he finishes successfully, or several users can train with the same agent successively.
- Emotional distance: Because culture and cultural communication is a quite critical theme, people might easily get offended when treated (in their opinion) wrongly. Additionally, trainees are often hesitant about trying novel nonverbal behavioral styles. Interacting with an agent, the user does not have to be afraid of doing something wrong or feeling embarrassment.
- Intensity: With a virtual agent, special nonverbal features can be displayed in varying intensities, allowing to highlight even subtle differences in behavior. An added benefit is the possibility of isolating certain features allowing the user to concentrate only on those features like, e.g. the spatial extent of a gesture.
- Generalization: The same agent and virtual scene can be used to simulate different cultures. Thus, the same system can be reused and adopted, for instance, to contrast the behavior of two cultures and point out the differences.
- Feedback: If the user's behavior is logged during an interaction, the agent can be used to replay this behavior and exemplify/emphasize problems or progress and can contrast the behavior either with previous behavior of the user or with the target behavior.

Although it is often claimed that virtual agents have positive effects on the learning experience, there are nearly no reliable large-scale evaluations so far that investigate the effects of experience-based role-plays with virtual characters in detail. One exception is the FearNot!v2 system. FearNot is an anti-bullying learning software that is designed to exemplify and let children test coping strategies for bullying in school in a safe environment. An evaluation study has been conducted with 1129 school children in two countries to evaluate the effects of employing virtual agents in training systems [37]. Whereas interaction in FearNot was purely text driven, a follow-up system has been introduced, which makes use of the same agent architecture and integrates also some non-verbal behaviors ([2, 29]).

The general idea behind experience-based role-plays is situated learning (e.g. [5, 42]). In this paradigm, learning has to take place in specific situations which provide rich contextual clues. Transferred to the language learning scenario for instance, instead of learning the dialogue for buying bread in class, you go to an actual bakery and buy bread there, i.e. try out your language knowledge in the right context. This of course is not possible in most cases because often a new language is learned out of context, i.e. not in the countries where they can be applied. Thus, role-plays with virtual characters are good substitutes for creating the right context and to experience and learn in specific situations.

2.4 Intercultural training

With GAME, we aim at providing the means to train gestures anytime anywhere in role-plays following suggestions by Hofstede [14], who describes three steps of intercultural training:

1. Awareness: The first step of gaining intercultural competence is being aware and accepting that there are differences in behavior. The hardest part of this learning step is to accept that there are no better or worse ways of behaving and especially that one's own behavior routines are not superior to others. To realize this step in a learning system with virtual agents, the trainee is confronted with a group of characters displaying the behavior routines of the target culture. With the knowledge of the trainee's cultural background, the agents could also contrast the behavior of the target culture with the behavior of the trainee's culture. Comparing the behavior patterns the trainee recognizes that there are differences but might not be able to pin them down.
2. Knowledge: In the second step, the trainee's knowledge of what exactly is different in the behavior is increased, which can be interpreted as getting an intellectual grasp on where and how one's own behavior differs. For instance, the trainee might have felt a little bit uncomfortable in step one due to a different pattern of gaze behavior. In step two, he will gain the knowledge on how his patterns differ from the patterns of the target culture and what the consequences are. In the learning system, the user is confronted with reactions to his behavior by his interlocutors. For instance, the agents could move away if the user comes too close. Moreover, the agents could replay specific behavior routines of the user and contrast them to the behavior routines of the target culture, pointing out where exactly the user's behavior deviates from the target culture.
3. Skills: Hofstede argues that the first two steps are sufficient to avoid most of the obvious blunders in intercultural communication. If the trainee has the ambition to blend into the target culture and adapt his

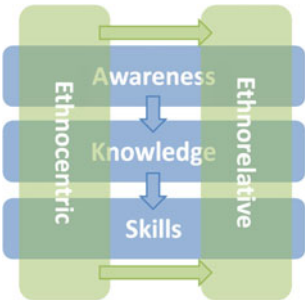

Fig. 1 Two dimensional model of intercultural coaching

own behavior, a third step is necessary: the training of specific nonverbal communication skills. If, e.g. avoiding eye contact in negotiations is interpreted as a sign of disinterest in the target culture, it might be a good idea to train sustained eye contact for such scenarios. Again, virtual characters can play a vital role in this learning step due to the aforementioned features.

Apart from the three steps introduced by Hofstede, Bennett [3] argues concisely that the success of a learning session is tightly related to the trainee's stage of intercultural awareness, which in general is ethnocentric at the beginning and with increasing awareness becomes more and more ethnorelative. He establishes a succession of six stages (three ethnocentric and three ethnorelative) that the trainee passes through and that differ in applicable teaching methods. On a conceptual level, the step from an ethnocentric to an ethnorelative perspective is essential in the development of intercultural competencies. Consequently, a full-blown contextual coaching application for cultural awareness will have to take all these dimensions into account by integrating the two-dimensional model depicted in Fig. 1.

With GAME, we present a first step in this direction. The system integrates interactive role-plays with virtual characters focusing currently on the knowledge and skills training for culture-specific gestures. To this end, a mobile serious game is realized where the user acquires knowledge about German emblematic gestures and then trains to perform these gestures in role-plays with virtual agents. A mobile platform was chosen for this approach because the ultimate goal is a coaching system that allows for contextual training sessions anytime and anywhere tailored to the user. The next section introduces the building blocks of the system focusing especially on the user interaction, i.e. the gesture recognition and the authoring of learning units.

3 The GAME approach: building blocks

Figure 2 gives an overview of the whole GAME architecture. GAME has been realized as a collaborative mobile environment for Window Mobile and tested on a HTC Touch Diamond. The user can choose to either run GAME in single-user mode or in competitive mode. Collaboration can be implemented locally by one user becoming the master, the others the slaves or remotely by connecting to the GAME server. The user can load new scenarios as well as gestures along with classifiers from the server. Content is authored by an XML-based authoring tool that allows specification of narrative structure, cut scenes and gesture information, and can be carried out by expert community members from the target culture.

By its experience-based role plays with virtual characters, GAME brings together ideas from situated learning and intercultural training in an integrated approach and paves the way for new m-learning concepts. Relating to the three steps of intercultural training by Hofstede (see Sect. 2.4), the game approach focuses on the second and third steps assuming that the user already has a certain level of cultural awareness. Thus, by playing with the system the user acquires knowledge and skills of culture-specific behavior, in our example about German emblematic gestures. To this end, the system features two modes, one dedicated to training specific skills (training mode, Sect. 4.1), the other allowing the user putting his new knowledge and skills to a test in specific situations like a visit to a beergarden (game mode, Sect. 4.2).

Thus, the current learning goal is training of emblematic gestures. According to McNeill [30], emblems are a special type of gestures. In general, gestures accompany speech and deliver either redundant or additional information about what was said in the utterance, e.g. using a pointing gesture to single out a specific referent that is mentioned in an utterance. Emblems, on the other hand, are not necessarily co-verbal but have a specific meaning in themselves. The American OK-sign is such an example. Emblems are also culture-specific in two ways. First, there are different sets of emblems in different cultures and second, the same emblematic gesture can have different meanings in different cultures. Consider again the American OK-sign, which is interpreted as an insult in Italy.

One of the main building blocks of GAME is the gesture recognition because this is the central interaction technique for the user. It is described next.

3.1 Mobile gesture recognition

Gesture-activated mobile edutainment aims at training German emblematic gestures. Thus, the user's gestural input has to be classified. Current smartphones offer acceleration sensors, which can be utilized to this end. Accelerometer-based gesture recognition has been shown to work at a high level of accuracy (e.g. [26, 33, 38, 41]). Based on previous work on gesture recognition with

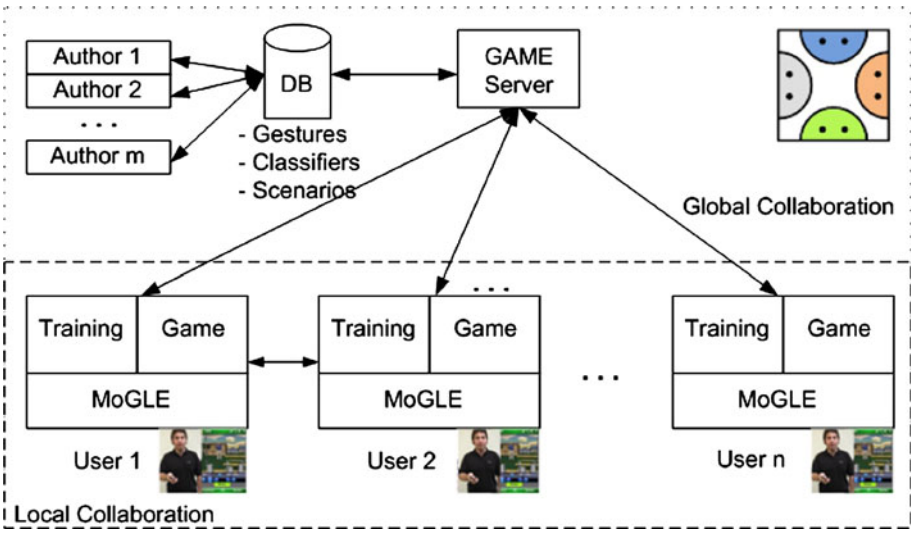

Fig. 2 The GAME architecture

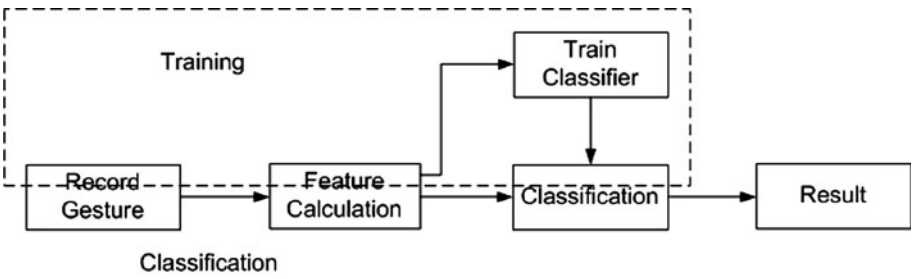

Fig. 3 The standard classification pipeline has been integrated in MoGLE

Nintendo's Wiimote controller presented in [33], we aimed at utilizing the acceleration sensors of handhelds for the same end. Thus, the general ideas from [33] have been adapted. In order to become leaner and faster to operate on the restricted environment of a mobile phone, Mobile gesture learning environment (MoGLE) restricts the number of available features and offers only a Naïve Bayes classifier in order to minimize calculation efforts on the mobile device.

Figure 3 illustrates the standard classification process that has been integrated in MoGLE. To train the classifier, a training set is recorded for each gesture class preferably by different users. Features are calculated on the raw signals and the resulting feature vector along with the information about the gesture class is used to train the Naïve Bayes classifier. For real-time classification, features are calculated for each gesture and the classifier calculates the most likely class for the feature vector. Currently, MoGLE is running under WindowsMobile on an HTC Touch Diamond. The acceleration sensors are working with a frame rate of 60 Hz for each axis. On the raw data, standardized statistical features are calculated for each axis: minimum, maximum, length, mean, median, and gradient.

Different evaluations were run to ensure that performance is comparable to the results presented earlier. In [33], we have shown that accelerometer-based gesture and expressivity recognition is robust and reliable.

To evaluate MoGLE, we replicated one of the experiments done with the Wiimote. The gesture set used as our benchmark is a set of control gestures for a video recording device, which were first introduced by Mäntyjärvi and colleagues ([20, 26]). Thus, using this gesture set allows us to evaluate MoGLE against two reference applications. In the original approach by Mäntyjärvi et al. [26], the raw acceleration data are quantified and then used for training hidden markov models (HMMs), i.e. no higher level feature calculation is done on the gestures. In principle, HMMs could be used for continuous gesture recognition, but the test set for the VCR control does not take this advantage into account rendering the original classification problem easily solvable by classification methods that require less computing power like Naïve Bayes. The VCR control gesture set is given in Fig. 4.

In [26], different training procedures have been tested in order to increase the recognition rate of the classifier. The best result that was achieved is 97.2% accuracy. This is taken as the benchmark to compare MoGLE against. Gestures were recorded under the same conditions. One user did 30 gestures per class, which were recorded in two sessions. In each session, 15 gestures per class were

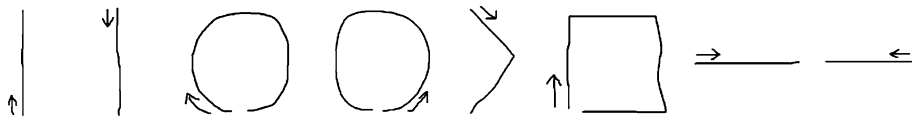

Fig. 4 VCR control gestures: *From left to right* gestures for play, stop, next, previous, increase, decrease, fast forward, fast rewind

Table 2 German emblems selected for GAME

Name	Gesture	Description
Come Here	Waving a hand rhythmically towards the body	Signaling a person to come closer
Go Away	Waving a hand rhythmically away from the body	Signaling a person to go away
Handshake	Moving right hand rhythmically up and down	Greeting someone
Go On	Rotating hand in front of body	Signaling a person to come to a conclusion
Unsure	Rotating one's hand back and forth (A23)	Signaling not being sure about a topic
Get Up	Raising upwards-pointing flat hands (A26)	Signaling a person to stand up
Eating	Putting hand to mouth	Asking for/Offering something to eat
Drinking	Drinking from a container (A05)	Asking for/Offering something to drink
Yummy	Rubbing splayed hand in circle across tummy	Signaling that food was good
Idiot	Pointing with index finger to forehead	Reproaching someone for being an idiot
Stupid	Waving a hand in front of one's eyes (A01)	Reproaching someone for being stupid
Threat	Cutting the throat (A21)	Threatening someone
Me	Pointing with index finger to own chest	Selecting oneself
No	Moving hand horizontally back and forth (A04)	Signaling disagreement
Time	Indicating to one's wrist (A02)	Indicating that time is running out, somebody is late

BLAG index in parentheses if applicable

performed. Recognition rates were calculated by a 14-fold cross-validation. The experiment was replicated for the Wiimote and showed that the faster, computationally less complex Naïve Bayes classifier is sufficient to solve the recognition task for a given user with a recognition rate of 99.6% for the eight-class problem [33] . For MoGLE, the gesturing device was changed from the Wiimote to a mobile device and running the classification process on the device itself produces comparable results with a recognition rate of 95.8%.

Having shown that the gestures are reliably recognizable, we aimed next at evaluating the performance of MoGLE for our task of German emblematic gestures. Fifteen emblematic gestures have been selected that are partly derived from the Berlin dictionary of German everyday gestures (Berliner Lexikon der Alltagsgesten, BLAG[1]) and partly based on their usefulness in the selected training scenarios (see Sect. 4.2). Table 2 gives an overview of the selected gestures along with their index in the BLAG (given in parentheses, if applicable) and a short description of their meaning.

[1] http://www.ims.uni-stuttgart.de/projekte/nite/BLAG/ (30 March 2011).

Performing gestures with the mobile phone might differ from a hands-free performance of the same gesture. To get insights into how users handle the device when performing each gesture, data were collected from a focus group of eight persons. Each person was asked to take the mobile phone and perform the gesture several times. Figure 5 gives some snapshots of the recordings for gesture "Go On". The information gathered from these tests was used to create the database of training samples for the classifier. To train the classifier, three trainers provided 10 training samples for each gesture resulting in a database of 450 gestures. Table 3 gives an overview of the results of a tenfold cross validation on this training database. The mean recognition result for the 15-class problem is 93.8%, which is a reasonable result for employing the classifier in the game and comparable to the results obtained earlier.

3.2 Enculturating virtual agents

The agents that serve as training partners have been enculturated making use of results from previous research [36]. The aim of this research was a model for adapting the interactive behavior of virtual agents to a given cultural background. To this end, a theory-driven top-down

Fig. 5 Snapshot from three users performing the "Go On" gesture with the mobile phone

Table 3 Recognition results for the fifteen emblematic gestures

Gesture	Rec. rate	Gesture	Rec. rate
Come Here	0.74	Yummy	0.97
Go Away	0.90	Idiot	0.92
Handshake	0.93	Stupid	0.95
Go On	0.98	Threat	0.98
Unsure	0.95	Me	0.97
Get Up	0.97	No	0.95
Eating	0.95	Time	0.95
Drinking	0.98	Average	0.94

approach was combined with an empirically driven bottom-up approach. The underlying theoretical model relies on Hofstede's idea of cultural dimensions [15] and exploits the assumed correlation between a culture's position on the dimensions and observable behavioral heuristics. Because those are only rough guidelines for actually generating appropriate behavior in an agent, face-to-face encounters in two different cultures (Germany and Japan) were analyzed to back up the model with empirical data. A Bayesian network model was developed that allowed inferring non-verbal behavior patterns if evidence was set for the cultural dimensions (details on the analysis and the model can be found in [36]). Cultural influences are apparent on all levels of the behavior planning and generation process. In [34], we have shown how the different levels of the network are exploited at different times of the behavior generation process, resulting in believable culture-specific behavior of the virtual agents. Figure 6 exemplifies this by a snapshot from German and Japanese face-to-face encounters and the generated behavior for the agents, in this case emphasizing differences in preferred postures.

Whereas in our previous work agents reacted in real-time to the user's input, the limited processing power of the mobile device makes it impossible to run the agent animation engine on it. Thus, we resorted to the solution of generating input-specific cut scenes by using the event flows of the scenarios to extract possible user interactions and simulate those with the original system. The result are interactive narratives for the scenarios that allow the users to explore all the possible paths through the scenario by his successful or unsuccessful attempts at gestural input. The next section details how scenarios are defined and thus how user input and cut scenes are specified.

3.3 Authoring of learning units

Figure 2 depicts the possibilities of training gestures and classifiers as well as authoring the content of the learning

Fig. 6 Differences in posture for German and Japanese samples (*left*) and generated behavior for agents (*right*)

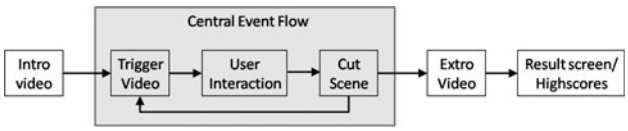

Fig. 7 Overview of general game flow with central interaction sequence highlighted

scenarios by expert community members. Based on [25], two learning scenarios have initially been realized taking into account different numbers of gestures. The "Greeting" scenario will serve as the example for detailing the authoring process. Figure 7 introduces the general game flow with the central interaction loop highlighted and Fig. 11 gives one example for the central interaction loop with a trigger video showing an agent waiting in the beergarden (left), the user performing the "Come Here" gesture (middle) resulting in a cut scene, where the agent moves towards the user (right). An XML structure allows to specify finite state machines with conditional transitions that evaluate the user's performance. Figure 8 gives a detail of the state machine for the "Greeting" scenario that deals with the sequence depicted in Fig. 11. Each video from the central event flow constitutes one state; the transitions correspond to the user interactions. In the example, the trigger video is the first state and shows an agent waiting in the beergarden. The user now performs a gesture and depending on his performance one of three successor states is activated. If the performance was really bad, i.e. the "Come Here" gesture was recognized with a probability of less than 0.5, the system remains in the state "Agent waiting". If the performance was good, i.e. recognition probability greater than 0.75, the system moves into the state "Agent moves to user" and the corresponding video of this cut scene is played. After that there is an unconditioned transition to the next trigger video that corresponds to state "Agent offers drink". If the user's performance is less than optimal but still acceptable, i.e. recognition probability between 0.5 and 0.75, the system moves to the state "Agent moves closer" and the corresponding video of this cut scene is played. This cut scene then serves also as the next trigger video, as the user has not yet succeeded in his task. In order to not frustrate the user by repeated failures, the thresholds for the evaluation of the next user gesture are relaxed somewhat in that a recognition probability of over 0.5 will be counted as a success.

The finite-state machine translates into a corresponding XML-structure that is depicted in Fig. 9. Along with the resources needed for the scenario like gestures and video files, the XML-structure specifies the flow of the interaction as well as the conditions for the transitions between states.

Two types of resources have to be specified for each scenario. The first resource are the gestures that are used in the scenario (<GESTURE>) along with the training samples necessary to train the classifier for this scenario (<TRAIN-DB>). That means that for each new scenario the classifier has to be trained based on the information from the scenario description. This modular approach allows tailoring the classifiers to the gestures used in the scenario, increasing the recognition rate. Moreover, it is easy to integrate new gestures as long as the training samples are provided along with the gesture names. The second resource are the movie files for the trigger and cut scenes. The names for the movies are specified by using the <MOVIE> tag.

The finite state machine is the second part of the specification and uses the <FSM> tag. The <SCENE> tag specifies the different states of the finite state machine. Transitions can either be conditional (<GESTURE-CHECK>) or unconditional (<GOTO>). If conditions are specified they can either be given by specifying the exact recognition probabilities or they can be given making use of some predefined values, which are employed in the example: `high` evaluates to greater than 0.75, `med` to greater than 0.5 but less than 0.75, and `low` to less than 0.5.

Apart from authoring the content of the system, it is possible to localize the interface because the idea is that the system should be used in a variety of target cultures. Localizing the interface is straightforward and currently

takes into account the texts used in the interface. All textual information in the system like button and menu labels as well as instruction texts is fully configurable without resorting to the source code. Labels and texts are read from external files during the startup phase and can be edited with any text editor.

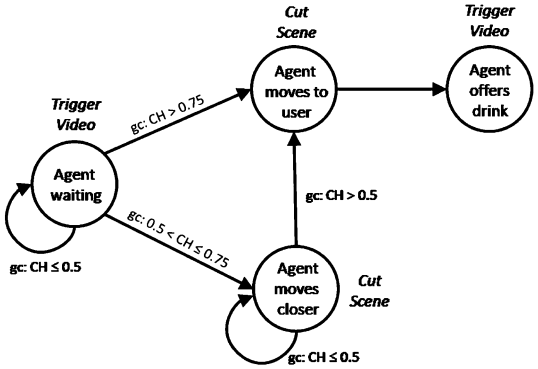

Fig. 8 Finite state machine for the example sequence of the "Greeting" scenario (*gc* gesture check, *CH* "Come Here" gesture)

4 The GAME approach: the application

4.1 Training mode

The application consists of two modi, a training mode to concentrate on specific gestures and the game mode for the experience-based role play. If the user chooses the training mode, he is able to acquire in-depth knowledge about single gestures as well as to practice his skills in doing the gestures. Thus, the training mode allows training and rehearsing in isolation without having to concentrate on the contextual factors for gesture use.

Fig. 9 XML-structure corresponding to the detail of the finite state machine from the "Greeting" scenario

```
<SCENARIO name="Beergarden" startscene="Intro">

<RESSOURCE URL="localhost/GAME/beergarden">
   <GESTURE name="Come Here">
      <TRAIN-DB count="30">ComeHereDB</TRAIN-DB>
   </GESTURE>
   <GESTURE name="Drink">
      <TRAIN-DB count="30">DrinkDB</TRAIN-DB>
   </GESTURE>
   <GESTURE name="Go On">
      <TRAIN-DB count="30">GoOnDB</TRAIN-DB>
   </GESTURE>
   ...
   <GESTURE name="Yummy">
      <TRAIN-DB count="30">YummyDB</TRAIN-DB>
   </GESTURE>
   <MOVIE>Intro</MOVIE>
   <MOVIE>Waiting</MOVIE>
   <MOVIE>Move2User</MOVIE>
   ...
   <MOVIE>Extro</MOVIE>
</RESSOURCE>

<FSM>
   ...
   <SCENE name="Waiting">
      <GESTURECHECK name="Come Here" high="Move2User"
                                     med="PartialMove"
                                     low="Waiting" />
   </SCENE>
   <SCENE name="Move2User">
      <GOTO name="OfferDrink" />
   </SCENE>
   <SCENE name="PartialMove">
      <GESTURECHECK name="Come Here" high="Move2User"
                                     med="Move2User"
                                     low="PartialMove"
      />
   </SCENE>
   ...
</FSM>

</SCENARIO>
```

Fig. 10 Training sequence for gestures in the "Greeting" scenario: gesture selection, information text, video sequence, gesture execution

Figure 10 gives an overview of the training cycle. The start screen (Fig. 10 left) offers three options to the user: (a) gesture training, (b) quick training, and (c) random training. The standard option is (a) gesture training. By selecting this option, an information text about the gesture is presented next, giving details about the meaning and usage of the gesture (Fig. 10 second from left). The user can now choose to directly try out the gesture (Button "Weiter"), or to see a small video of how the gesture is performed (Button "Video"). A snapshot from such a video is given in Fig. 10 (second from right). Having seen the video, the user now performs the gesture and gets the feedback on his performance in auditory and textual form. The recognition result for the user's performance is shown in Fig. 10 (right), where the user did not perform very well as the system did recognize completely different gestures ("Go On" with a probability of around 70% and "Yummy" with around 30%). The gesture is performed by pressing on the gray area (e.g. with the thumb), where also the recognition results are displayed, and releasing this press after the gesture has been performed. After each gesture performance the recognition results are given in auditory and textual form. This can be repeated until the user is satisfied with the result.

If the user chose (b)—quick training—instead of the standard gesture training at the beginning, he jumps directly to the gesture execution without information on the gesture and how it is performed. If necessary the information text as well as the video can be requested at any time by pressing the "Info" and "VID" buttons, respectively (Fig. 10 right).

The last option (c)—random training—allows the user to rehearse what he has trained before by presenting a random gesture from the list of available gestures, which the user has to perform. This mode was integrated for motivational reasons to keep the training session more engaging.

4.2 Game mode

The game mode realizes the experience-based role play and is based on standard techniques for intercultural training [25]. Two scenarios have been integrated so far: "The Greeting" and "The Visit". The greeting allows the rehearsal of greeting rituals in the target culture, whereas the visit represents a less formal interaction during dinner with a family in the target culture. In GAME, both scenarios take place in a beergarden (typical Bavarian meeting place) and differ in length and number of gestures that are performed (5 during the greeting, 10 during the visit).

The original greeting scenario is generally situated at an airport, train station or similar location, where the trainee arrives and is met by a host from the target culture. What follows is like the first chapter of each language textbook augmented by the appropriate non-verbal behavior. The host will first welcome the trainee followed by a self-introduction and some questions on the setting, i. e. the journey. The trainee applies his verbal and non-verbal skills of the target culture and comes up with the right phrases and behavior (e.g. performing a handshake). The scenario has been adapted to the beergarden environment and focuses solely on the gestural interaction. The user is identified by an agent and has to perform the right gesture, in this case a wave to greet the agent followed by a signal to come over to the user. The agent then moves to the user and proposes to drink something together, which the user accepts by repeating the drinking gesture. Both then comment on the quality of the beverage with the yummy gesture before parting again performing a handshake.

The original visit scenario takes place at the host's private home. The trainee has been invited for dinner and now encounters the host's family during this social event. Often, this scenario includes moments of conflict and tension, when, e.g. an inappropriate small talk topic is chosen. The visit scenario has been adapted to the beergarden

environment. It follows the greeting scenario until user and agent share a drink. To include the moment of conflict and tension, a second (drunk) agent now enters the scene which starts insulting the first agent. The user's task is to apply some of the gestures to get rid of the second agent, i.e. the idiot and go away gestures. Afterwards, common ground is re-established with the first agent by commenting on the conflict by means of the stupid gesture and inviting the agent for another drink. Both then comment on the quality of the beverage before signaling that it is time to move on. They part performing a handshake gesture.

Especially the greeting scenario is important in intercultural encounters as one of the most basic interaction rituals. It has been argued that such a first encounter serves several social aspects of establishing common ground in a safe and face-keeping manner (e.g. [1, 21, 40]). Thus, first meetings are always of ritualistic nature where the script is highly culture-specific.

In GAME, both scenarios take place in a beergarden and are technically realized as interactive narratives. A short video is presented that triggers a reaction of the user in the form of a gesture. Depending on the gesture and its performance a cut scene is played, which in turn leads to another trigger video. To give a short example (Fig. 11), the greeting scenario starts with the user entering the beergarden and noticing an agent that is apparently waiting for someone. The user's reaction should now be to either wave hello or signal the agent to come closer. The latter will for instance result in a video showing the agent moving closer to the user.

The scenarios force the user to apply his knowledge about the culture-specific emblematic gestures in the context of their use, thus realizing a simulated situated learning experience.

5 Evaluating GAME

Two different types of evaluations were conducted:

1. Usability evaluation
2. Evaluation of training effect

The first type focused on the general usability of the system, exploring its hedonistic and pragmatic qualities. The second type focused on the claim that the experienced based learning with virtual characters can help improve skills training.

5.1 Usability evaluation

In order to show that the resulting interface and the game play are attractive to users, an exploratory evaluation was conducted on a public event for the German year of science in 2009 that took place in the city center of Augsburg. For this event, the Department of Computer Science presented a number of interactive demos along with information on the study programs. During this event, participants were recruited on site.

5.1.1 Design

20 participants could be won (15 male, 5 female) for the study, which consisted of a training phase followed by a single player role play with the greeting scenario. Afterwards participants filled out an AttrakDiff questionnaire [13], which is used to measure the hedonistic and pragmatic qualities of the system. Additionally, participants were asked to give their subjective impressions about the input possibilities and the game play. Thus, three different sources of information are available for the evaluation:

1. Log data: All user actions have been logged during training phase and role play allowing analyzing the success of gesture executions.
2. Hedonistic and pragmatic quality: By requesting a graded response to adjective pairs like "complicated–simple", the AttrakDiff questionnaire results in a rating of the product's hedonistic and pragmatic qualities.

Fig. 11 A short game sequence with the user reacting to a waiting agent that moves closer if the gesture is performed correctly

3. Subjective impressions: Participants have been asked to write down their subjective impressions about the game play and the gestural input possibilities.

5.1.2 Results and discussion

5.1.2.1 Log data In this explorative analysis we wanted to find out if users are able to handle the device and successfully play the game by performing gestures and if the training mode has an effect on the gesture performance in the game. For the analysis we divided the users into low performers with success rates below 0.5 and high performers with success rates above 0.5. The log data revealed that 7 of the 20 participants were low performers. Next, we compared the number of training rounds low and high performers did and saw that the low performers either directly started with the game or did on average less training rounds than the high performers. Figure 12 gives the box plot for this relation. What is apparent from the plot is that users with high success rates had on average more training rounds than users with low success rates. A correlation analysis (Pearson) showed a significant positive correlation (0.509, $p < 0.05$) between training and the success rates in the game.

Thus, the log data analysis highlights that although we designed the gestures and classifiers based on user observations, there is still a need for getting acquainted with handling the device to perform conversational gestures. This does not come as a complete surprise as users have never done this before. The time needed for trying out the device is not overly long because on average users need 7 training units for 5 gestures to become a high performer, i.e. basically they have to try out each gesture ones. On the other hand this result raises the question if training with the device will carry over to performing the gestures without the device. Evaluating this training effect is the topic of the Sect. 5.2.

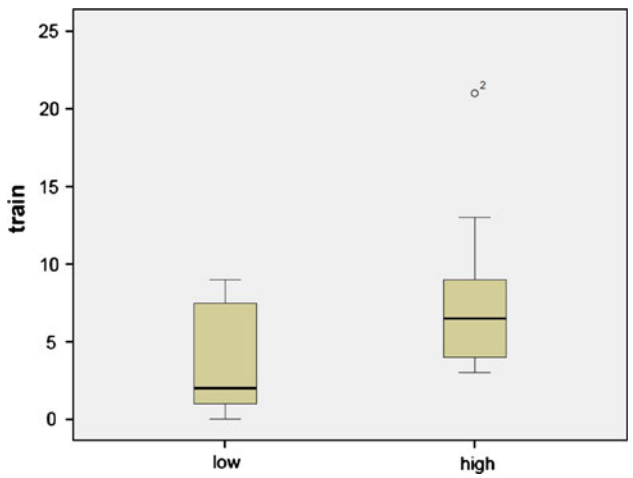

Fig. 12 Relation between number of training rounds and success rate

5.1.2.2 Hedonistic and pragmatic quality The AttrakDiff questionnaire asked the participants to select a graded response (seven point scale) to adjective pairs that fall into four different categories. Participants had to rate 28 pairs in all, i.e. 7 pairs for each category.

- Pragmatic quality (PQ): Describes the usability of the product and clarifies if the user can reach his goals with the system. An example pair for this category is "complicated–simple".
- Hedonistic quality-identity (HQ-I): Describes if the user is drawn into the interaction and can identify with the system. An example pair for this category is "unprofessional–professional".
- Hedonistic quality-stimulation (HQ-S): Describes if the product is stimulating in presenting new, innovative and motivating ways of interaction and content presentation. An example pair for this category is "conservative–innovative".
- Attractivity (ATT): Describes a global rating based on perceived quality of the product. An example pair for this category is "discouraging–motivating".

Figures 13 and 14 give the result of the AttrakDiff analysis. An overview for the hedonistic and pragmatic quality of the system is given in Fig. 13. It shows that users reacted positively towards the system on both dimensions, rating it as attractive to use and self-oriented, which means that the interaction was perceived as a positive experience for personal development. This result is compatible with the goals we had for the system because it was designed to support the user in his self-directed study of knowledge and skills of non-verbal behavior. The detailed analysis (Fig. 14) gives the mean ratings of all adjective pairs and corroborates the first impression. For nearly all pairs, the ratings are on the positive side. For two pairs (technical–human, unpredictable–predictable) results are rather neutral instead.

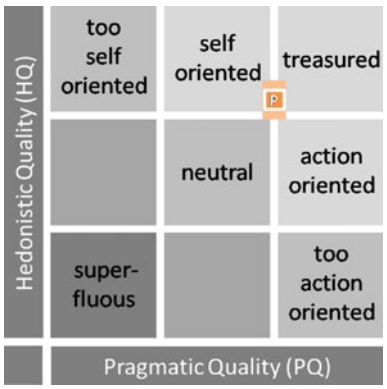

Fig. 13 Result of AttrakDiff evaluation (overview)

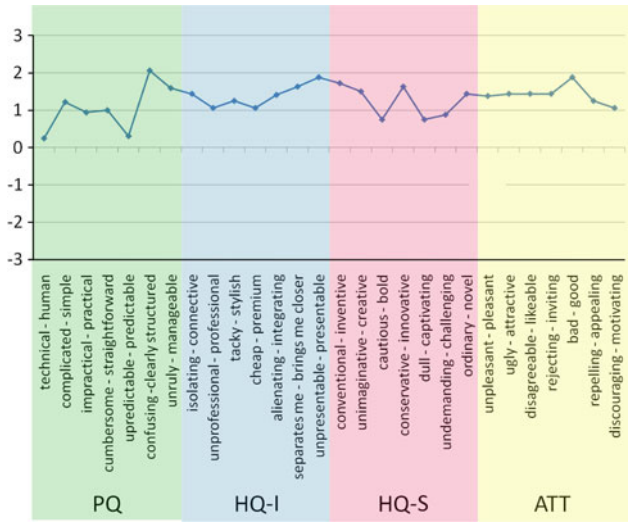

Fig. 14 Result of AttrakDiff evaluation (details)

Concerning the "unpredictable" versus "predictable" dimension, we observed that for the low performers it was not always clear why the system did not register their gestures as correct resulting in low ratings for this dimension because for them the system seemed to recognize their gestures on a random basis. A reason for the low score on the dimension "technical" versus "human" could be that conversational gestures are generally done without technical requisites. Thus, the gestural interaction becomes suddenly mediated by the mobile device, which introduces a technical layer to the interaction. For other gesture types this might not pose a problem, e.g. conducting an orchestra, which is often mediated by a baton. Moreover, the advent of game consoles that make use of acceleration sensing to introduce embodiment into the game play might also have an influence on this rating when users get more acquainted with gesture recognition devices.

5.1.2.3 Subjective impressions Consistent with the AttrakDiff results, users were quite positive about the interaction possibilities offered by the system and the game play. Two comments recurrently came up that should be considered during the further development. Some of the buttons were perceived as being too small, especially if the user did not use a stylus but operated the system solely with his fingers. The second comment concerned the event flow during the training mode. To select a new training gesture, the user always has to go back to the main menu (see Fig. 10 left). Several users requested a possibility to change the training gesture directly from the result screen (Fig. 10 right), for instance by introducing a next button.

The usability evaluation revealed the positive potential of our approach. Participants were able to handle the device and interact with the application successfully by performing gestures. The analysis of the hedonistic and pragmatic qualities showed that the system is perceived as motivating and innovative by the users. The logical next step is to evaluate if the experience-based training has an effect on the user apart from being motivating.

5.2 Evaluation of cultural training

The usability study presented in the last section showed that success rates in the experience-based role-plays increase with acquaintance on handling the device. Thus, can there be a learning effect of the emblem training that carries over to doing gestures without the device? The evaluation presented in this section was conducted to test this assumption. It is based on suggestions by Elfenbein and Ambady [7] about taking the implicit cultural background of the participants into account as an experimental condition. Thus, the study is done in two steps, distinguishing between gesture performance in the game and perception of gesture performance by German participants.

1. Skills training: Participants from other cultures than the target culture interact with the system (test group) or learn about German emblematic gestures in a traditional way (control group).
2. Performance rating: Video recordings of gestures from the two groups are rated by participants from the target culture based on their implicit knowledge about good performance of the gestures.

5.2.1 Design

5.2.1.1 Step 1: skills training For the first step of the experiment, newly arrived Erasmus students have been recruited. 15 students participated in the study, all females, with an age ranging from 20 to 24 (mean 22.8). None of the participants have been in Germany before, but all were familiar with the language, which they learned in courses in their home countries. Ten participants were randomly assigned to the test group and five were randomly assigned to the control group.

The test group (TG) used the system to train the five gestures of the greeting scenario and employed them in the scenario. Afterwards, they were asked to perform the gestures without the mobile device and were videotaped during this performance. The control group (CG) used instead traditional textual descriptions of the gestures accompanied by still images following what is found in standard training material (e.g. [6, 31]). An example is given in Fig. 15. Afterwards, they were asked to perform the gestures and were videotaped during this performance. Then, the CG participants had the possibility to test out the system (training and play).

Fig. 15 Example of training material for control group. [Translation (by the authors) of the German gesture explanation: this gesture signifies that the mental health of the interaction partner is in question. The gesture is often accompanied by an appropriate facial expression]

Thus, the interaction sequences for the two groups were

TG training (training and play), recording of gesture performance
CG training (traditional), recording of gesture performance, system test (training and play)

The main goal of this first step was to gather the video material that is then rated by German observers in the second step.

5.2.1.2 Step 2: performance rating To assess if users performed better after training with the GAME system, the gesture performance was rated in a web-based study by German participants. 42 people participated in this study, 20 males and 22 females between the age of 20 and 43 with a mean of 28.7.

Each participant had to rate ten videos, i.e. one sample of each gesture from each group (test and control). The samples were randomly chosen from the available videos and presented in a random order. Figure 16 shows a screenshot from the study and Fig. 17 snapshots from two of the performance videos. Participants could watch the video as often as they liked. They were asked to write down the meaning of the gesture and additionally rate the quality of the gesture performance on a seven-point Likert scale. Moreover, participants were asked to indicate four performance features (speed, spatial extent, power, and fluidity) on a seven-point Likert scale.

The hypothesis of this evaluation considers the differences in performance for the two groups (test vs. control):

TG outperforms CG, i.e. performance ratings from native speakers are better for the test group and thus the experience based training results in a positive effect.

5.2.2 Results

Table 4 gives the results for the performance rating by the German native speakers.

Gestural expressivity has not been analyzed yet, but will be compared with the results obtained in an earlier study [36]. Results from Table 4 give the mean rating for each gesture from the German participants for the test and control groups. A t test was run on the data and reveals highly significant differences for three of the gestures (Come Here, Stupid, Go On; $p < 0.01$ for each gesture), with higher ratings for the test group, i.e. those participants that trained with the GAME system. For the other two gestures (Eating, Drinking), no significant difference in performance could be seen by the German observers.

5.2.3 Discussion

Regarding the hypothesis, the results are partly supporting it. For three gestures (Come Here, Stupid, Go On) the experience-based approach works significantly better, and for the remaining two (Eating, Drinking) there is no difference between the test and control groups. For the drinking gesture, the experience-based approach was even counterproductive, leading to reduced ratings relative to the control group although the difference is not statistically significant. There is one obvious reason for the result concerning the drinking gesture and that is the obtrusiveness and unintuitive handling of the phone for this gesture. While designing the gesture, the main idea was to use the phone as a kind of container from which the user is drinking thus having a natural way of realizing this gesture. Although that was helpful for the game scenarios, the movement did not translate properly to the case where the device was no longer present. For the eating gesture it was our impression from the comments of the participants that this was a widely known gesture by the participants and thus not a good choice as a test case.

We draw the following conclusions from the results of this evaluation. It is crucial to carefully design the training scenarios to include gestures that are suitable for training in an experience-based manner making use of a recognition device that is actively handled by the user. Some gestures

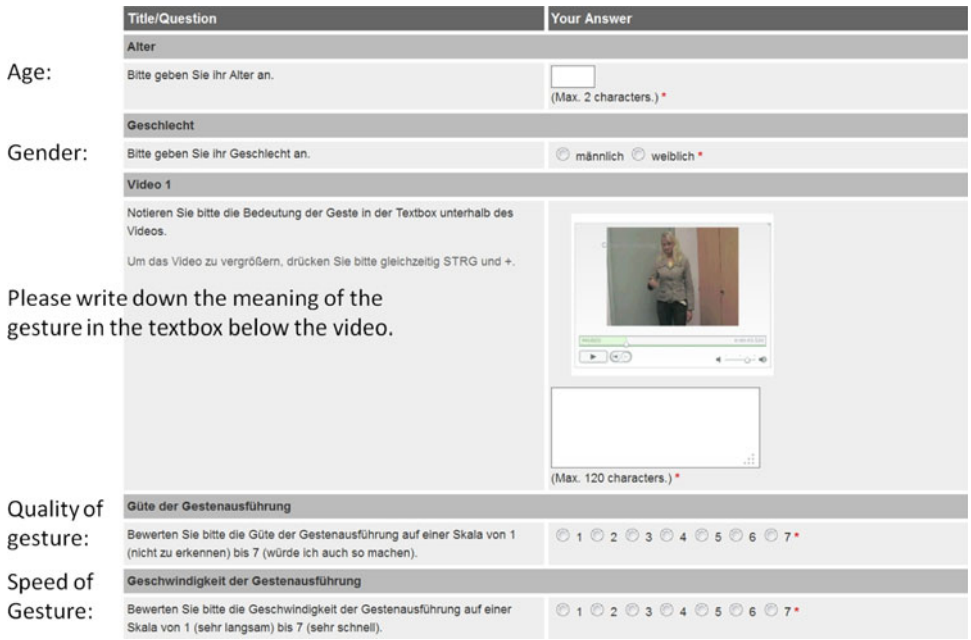

Fig. 16 Screenshot from web form (in German)

Fig. 17 Examples from performance videos for gesture "Go On". *Above* participant from test group. *Below* participant from control group

Table 4 Results from performance rating (*t* test)

Gesture	Test group	Control group	*p* value
Come Here	5.71	3.45	0.00
Eating	4.48	4.19	0.26
Drinking	4.64	5.10	0.12
Stupid	6.33	2.29	0.00
Go On	3.07	1.24	0.00

are not suitable to be trained with such device as very unnatural movements or movements that do not translate to the non-device case can emerge. For the gestures that are suitable, the evaluation was a great success with a rating of the gesture performance that was significantly better than with the traditional method.

A further analysis of the results reveals that there seem to be "hard" gestures for which the movement is not easy to grasp but that nevertheless profit from the use of the experience-based approach. This refers to the Go On gesture that was rated significantly better with the experience-based approach but still was rated below average.

The first step of the evaluation revealed another interesting result that is worth pursuing further. Both groups, i.e. test and control, played the game scenario "The Visit", the test group as part of their gesture training, the control group after the training sessions, and the recording of the gesture performance. What is evident from the log files is that CG outperforms TG in terms of successful scenario interactions. Thus, the mix of different materials and the repetition seem to be beneficial to CG for employing the gestures in a concrete scenario. The overall conclusion from the results is that the experience-based training has

great potential as a means to try out and train gestural performance, i.e. to serve as knowledge and skills training.

6 Conclusion

The work presented in this paper is based on the idea of marrying mobile technology with the possibilities of experience-based role-plays to support a coaching approach. It draws its motivation from two sources. First, virtual characters have been shown to be successful tools for intelligent tutoring systems. Second, intercultural training is facing a shift towards coaching endeavors. With GAME we presented a first step in this direction. A mobile edutainment platform has been developed that challenges the user with active tasks where he has to put his knowledge and skills about non-verbal behavior to a test in interactions with virtual characters. To this end, the GAME platform offers gesture recognition and authoring possibilities. Scenarios are defined as finite state machines with conditioned transitions between states. Two evaluation studies have been presented that show the positive potential of this approach and highlight the fact that the experience-based gesture training outperforms traditional methods.

So far, the experience-based role-plays with virtual characters have been brought to the mobile device, freeing the user from desktop-based stationary interactions. The aim is to realize a coaching approach that takes the user's context (location, agenda, etc.) into account for suggesting a learning session. Thus, a proactive system is envisioned as the next step that decides on scenarios based on contextual clues like location or the user's agenda. Ideally, it should also take the user's stage of intercultural development (ethnocentric to ethnorelative) into account.

References

1. Argyle, M.: Bodily Communication. Methuen, London (1975)
2. Aylett, R., Paiva, A., Vannini, N., Enz, S., André, E., Hall, L.: But that was in another country: agents and intercultural empathy. In: Proceedings of 8th International Conferecne on Autonomous Agents and Multiagent Systems (AAMAS), pp. 329–336 (2009)
3. Bennett, M.J.: A developmental approach to training for intercultural sensitivity. Int. J. Intercult. Relat. **10**(2), 179–195 (1986)
4. Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., Yan, H.: Designing embodied conversational agents. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds) Embodied Conversational Agents, pp. 29–63. MIT Press, Cambridge (2000)
5. Clancey, W.J.: A tutorial on situated learning. In: Proceedings of Computers and Education (1995)
6. Clayton, P.: Body Language at Work: Read Signs and Make the Right Moves. Hamlyn, London (2003)
7. Elfenbein, H.A., Ambady, N.A.: When familiarity breeds accuracy: cultural exposure and facial emotion recognition. J. Pers. Social Psychol. **85**(2), 276–290 (2003)
8. Endrass, B., Rehm, M., André, E.: Planning smalltalk behavior with cultural influences for multiagent systems. Comput. Speech Lang. **25**(2), 158–174 (2011)
9. Fowler, S.M., Blohm, J.M.: An analysis of methods for intercultural training. In: Landis, D., Bennett, J.M., Bennett, M.J. (eds) Handbook of Intercultural Training, pp. 37–84. Sage Publications Inc., Thousand Oaks (2004)
10. Hall, E.T.: The Silent Language. Doubleday, Garden City (1959)
11. Hall, E.T.: The Hidden Dimension. Doubleday, Garden City (1966)
12. Hall, E.T.: Beyond Culture. Doubleday, Garden City (1976)
13. Hassenzahl, M.: The effect of perceived hedonic quality on product appealingness. Int. J. Human Comput. Interact. **13**(4), 481–499 (2001)
14. Hofstede, G.: Cultures and Organisations—Intercultural Cooperation and its Importance for Survival, Software of the Mind. Profile Books, New York (1991)
15. Hofstede, G.: Cultures Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications, Thousand Oaks (2001)
16. Hofstede, G.J., Pedersen, P.B., Hofstede, G.: Exploring Culture: Exercises, Stories, and Synthetic Cultures. Intercultural Press, Yarmouth (2002)
17. Isbister, K.: Building bridges through the unspoken: embodied agents to facilitate intercultural communication. In: Payr, S., Trappl, R. (eds) Agent Culture: Human-Agent Interaction in a Multicultural World., pp. 233–244. Lawrence Erlbaum Associates, London (2004)
18. Jan, D., Herrera, D., Martinovski, B., Novick, D., Traum, D. et al.: A computational model of culture-specific conversational behavior. In: Pelachaud, C. (ed.) Intelligent Virtual Agents (IVA'07), pp. 45–56. Springer, Berlin (2007)
19. Johnson, W.L., Friedland, L.E.: Integrating cross-cultural decision making skills into military training. In: Schmorrow, D., Nicholson, D. (eds) Advances in Cross-Cultural Decision Making, CRC Press, Boca Raton (2010)
20. Kela, J., Korpipaää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., Marca, S.D.: Accelerometer-based gesture control for a design environment. Pers. Ubiquitous Comput. **10**, 285–299 (2006)
21. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge Univ Press, Cambridge (1991)
22. Khaled, R., Barr, P., Biddle, R., Fischer, R., Noble, J.: Game design strategies for collectivist persuasion. In: Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games, pp. 31–38 (2009)
23. Kim, J.M., Hill, R.W., Durlach, P.J., Lane, H.C., Forbell, E., Core, M., Marsella, S., Pynadath, D., Hart, J.: BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. Int. J. Artif. Intell. Educ. **19**(3), 289–308 (2009)
24. Kluckhohn, F., Strodtbeck, F.: Variations in Value Orientations. Row, Peterson, New York (1961)
25. Losche, H.: Interkulturelle Kommunikation. Sammlung praktischer Spiele und Übungen. Ziel, Augsburg (2005)
26. Mäntyjärvi, J., Kela, J., Korpipää, P., Kallio, S.: Enabling fast and effortless customisation in accelerometer based gesture interaction. In: Proceedings of MUM'04, pp. 25–31 (2004)
27. Marcus, A., Hamoodi, S.: The impact of culture on the design of arabic websites. In: IDGD'09 Proceedings of the 3rd International Conference on Internationalization, Design and Global Development, pp. 386–394. Springer, Berlin (2009)
28. Marcus, A., Krishnamurthi, N.: Cross-cultural analysis of social network services in japan, korea, and the usa. In: IDGD'09

Proceedings of the 3rd International Conference on Internationalization, Design and Global Development, pp. 59–68. Springer, Berlin (2009)
29. Mascarenhas, S., Dias, J., Prada, R., Paiva, A.: A dimensional model for cultural behavior in virtual agents. Appl. Artif. Intell. **24**, 552–574 (2010)
30. McNeill, D.: Hand and Mind—What Gestures Reveal about Thought. The University of Chicago Press, Chicago (1992)
31. Pease, A., Pease, B.: The Definitive Book of Body Language: How to Read Others Attitudes by Their Gestures. Orion, Houston (2005)
32. Rehm, M.: Developing enculturated agents—pitfalls and strategies. In: Blanchard, E.G., Allard, D. (eds) Handbook of Research on Culturally-Aware Information Technology., IGI Global, Hershey (2010)
33. Rehm, M., Bee, N., André, E.: Wave like an Egyptian—Acceleration based gesture recognition for culture-specific interactions. In: Proceedings of HCI 2008 Culture, Creativity, Interaction, pp. 13–22 (2008)
34. Rehm, M., Nakano, Y., André, E., Nishida, T. et al.: Culture-specific first meeting encounters between virtual agents. In: Prendinger, H. (eds) Intelligent Virtual Agents., Springer, Berlin (2008)
35. Rehm, M., Nakano, Y., André, E., Nishida, T.: Editorial for special issue on enculturating human computer interaction. AI Soc. **24**(3), 209–211 (2009)
36. Rehm, M., Nakano, Y., André, E., Nishida, T., Bee, N., Endrass, B., Wissner, M., Lipi, A.A., Huang, H.H.: From Observation to Simulation—Generating Culture Specific Behavior for Interactive Systems. AI Soc. **24**, 267–280 (2009)
37. Sapouna, M., Wolke, D., Vannini, N., Watson, S., Woods, S., Schneider, W., Enz, S., Hall, L., Paiva, A., André, E., Dautenhahn, K., Aylett, R.: Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. J. Child Psychol. Psychiatr. **51**(1), 104–112 (2010)
38. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a wii controller. In: Proceedings of Tangible and Embedded Interaction (TEI) (2008)
39. Schwartz, S.H., Sagiv, L.: Identifying culture-specifics in the content and structure of values. J. Cross-Cultural Psychol. **26**(1), 92–116 (1995)
40. Ting-Toomey, S.: Communicating Across Cultures. The Guilford Press, New York (1999)
41. Urban, M., Bajcsy, P., Kooper, R., Lementec, J.C.: Recognition of arm gestures using multiple orientation sensors: Repeatability assessment. In: IEEE Intelligent Transportation Systems Conference, pp. 553–558 (2004)
42. Vygotsky, L.S.: Thinking and speech. In: The Collected Works of L. S. Vygotsky, Problems of general psychology, vol. 1, pp. 39–285. Plenum Press, New York (1987)
43. Warren, R., Diller, D.E., Leung, A., Ferguson, W., Sutton, J.L.: Simulating scenarios for research on culture and cognition using a commercial role-play game. In: Kuhl, M.E., Steiger, N.M., Armstrong, F.B., Joines, J.A. (eds.) Proceedings of the 2005 Winter Simulation Conference (2005)